Behav Res (2016) 48:243–258
DOI 10.3758/s13428-015-0568-1

Trying to separate the wheat from the chaff: Construct- and faking-related variance on the Implicit Association Test (IAT)

Jessica Röhner · Torsten Ewers

Published online: 21 February 2015

Abstract Recent research has indicated that diffusion model analyses allow the user to decompose the *traditional IAT effect* (*D* measure) into three *newly developed IAT effects*: IAT_v, which has already been shown to be significantly related to the construct-related variance of the IAT effect, and IAT_a and IAT_{t0}, both of which have been assumed to provide an indication of faking. But research on the impacts of faking on IAT_v, IAT_a, and IAT_{t0} is still warranted. By reanalyzing a data set containing both faked and unfaked IAT effects, we investigated whether diffusion model analyses could be used to separate construct-related variance from faking-related variance on the IAT. Our results revealed that this separation is not yet possible. As had already been shown for the traditional IAT effect, IAT_v was affected by faking. Interestingly, it was affected by faking only under more difficult faking conditions (i.e., when participants were asked to fake without being given recommended strategies for how to do so, and when they were requested to fake high scores). By contrast, IAT_a was affected by faking only in the comparably easy faking condition (i.e., when participants had been informed about possible faking strategies and were asked to fake low scores). IAT_{t0} was not affected by faking at all. Our results show that although diffusion model analyses cannot yet provide a clear separation between construct- and faking-related variance, they allow us to peer into the black box of the faking process itself, and thus provide a useful tool for analyzing and interpreting IAT scores.

J. Röhner · T. Ewers
Department of Psychology, Chemnitz University of Technology, Chemnitz, Germany

J. Röhner (✉)
Department of Psychology, Chemnitz University of Technology, D-09107 Chemnitz, Germany
e-mail: jessica.roehner@psychologie.tu-chemnitz.de

Keywords Implicit Association Test (IAT) · Diffusion model analyses · Construct-related variance · Faking-related variance · IAT effects · Faking process

The Implicit Association Test (IAT; Greenwald, McGhee, & Schwartz, 1998) has attracted an enormous amount of research interest in recent years. It was designed to assess automatic implicit associations between two target concepts and an attribute dimension by using participants' reaction times. On the IAT, participants have to sort stimuli that appear consecutively in the middle of the computer screen into four different categories: (a)two contrasted target concept categories that form the target dimension and (b)two contrasted attribute categories that form the attribute dimension.

The IAT procedure using the extraversion IAT as an example

On an extraversion IAT (Back, Schmukle, & Egloff, 2009), as used in the present reanalysis, the target dimension includes self-relevant versus non-self-relevant words (e.g., me vs. others), and the attribute dimension includes extraversion-related versus introversion-related words (e.g., talkative vs. shy). The IAT consists of seven blocks altogether, of which Blocks 1, 2, and 5 are the so-called *single* or *practice* blocks, which introduce the target or attribute discrimination. In these blocks, the categories of either the target concepts or the attribute concepts are presented in the upper corners of each side (i.e., left and right) of the display screen. Participants are instructed to respond to exemplars of each category by pressing a key on the same side as the label. Blocks 3 and 4 as well as 6 and 7 are the so-called *combined* blocks, in which the attribute discrimination is paired with the target discrimination

(i.e., participants must assign words from all four categories in these blocks). Thus, on the extraversion IAT, in Blocks 3 and 4 (the compatible phase), participants must respond to self-relevant and extraversion-related words with one key and to non-self-relevant and introversion-related words with the other key. In Blocks 6 and 7 (the incompatible phase), participants must respond to introversion-related and self-relevant words with one key and to the extraversion-related and non-self-relevant words with the other key.[1]

The traditional IAT effect

The rationale behind the IAT is that the sorting task should be easier and thus completed more quickly when the two concepts that share one response key are strongly associated. If two concepts are only weakly associated, sorting them into one category should be more difficult and should therefore be conducted more slowly. The *traditional IAT effect* (i.e., the so-called *D* measure; see Greenwald, Nosek, & Banaji, 2003a, 2003b) is computed as the difference in reaction times between the incompatible phase and the compatible phase divided by their overall standard deviation. It is used as an indicator of the strength of the association between the concepts (e.g., self and extraversion for the extraversion IAT).

The problem associated with the traditional IAT effect

Although the validity of the IAT for measuring automatic associations has been documented in a number of studies (e.g., Banse, Seise, & Zerbes, 2001; Bar-Anan & Nosek, 2014; Gawronski, 2002; Greenwald et al., 1998; Hofmann, Gawronski, Gschwendner, Le, & Schmitt, 2005), the traditional IAT effect is associated with a relevant problem. It has been criticized for containing not only variance related to the construct but also method-specific variance (e.g., Back, Schmukle, & Egloff, 2005; McFarland & Crouch, 2002; Mierke & Klauer, 2003) and faking-related variance (e.g., De Houwer, Beckers, & Moors, 2007; Fiedler & Bluemke, 2005; McDaniel, Beier, Perkins, Goggin, & Frankel, 2009; Röhner, Schröder-Abé, & Schütz, 2011, 2013; Steffens, 2004). Unfortunately, using the traditional IAT effect (i.e., the *D* measure; see Greenwald et al., 2003a, b) does not allow the researcher to separate these sources of variance from each other. In other words, all variance will usually be treated as construct-related variance.

[1] The presentation of the combined phases can be counterbalanced in IATs. Hence, as a researcher, one can decide whether the participant will work on the compatible phase first and afterward on the incompatible one (i.e., the sequence of the IAT explained above), or whether the phases should be presented the other way around. In order to avoid unnecessarily complicating the description of the IAT, only the most common order is presented here.

However, recent research by Klauer, Voss, Schmitz, and Teige-Mocigemba (2007) has indicated that so-called *diffusion model analyses* allow the user to decompose the traditional IAT effect (i.e., the *D* measure; Greenwald et al., 2003a, b), and thus may be useful for separating these different sources of variance from each other.

The diffusion model

The diffusion model (Ratcliff, 1978, 2014; Ratcliff, Gomez, & McKoon, 2004; Ratcliff & Rouder, 1998, 2000) represents a stochastic model for binary decision tasks. It enables researchers to estimate several parameters associated with people's decision processes. Accordingly, it has been successfully applied to data from a variety of decision tasks such as recognition memory (e.g., Spaniol, Madden, & Voss, 2006), lexical decisions (Ratcliff, Gomez, et al., 2004; Ratcliff, Thapar, Gomez, & McKoon, 2004), perceptual discrimination (Voss, Rothermund, & Brandstädter, 2008; Voss, Rothermund, & Voss, 2004), priming effects (Voss, Rothermund, Gast, & Wentura, 2013), the IAT (Klauer et al., 2007; Schmitz & Voss, 2012; van Ravenzwaaij, van der Maas, & Wagenmakers, 2011) and others (for a review, see Wagenmakers, 2009).

The diffusion model is built on the assumption that people continuously accumulate information (i.e., response evidence) from the stimuli presented in a task in order to make their decisions. After collecting sufficient information, they make their decision.

Here, we will explain the process underlying the diffusion model by visually representing a sample decision path for one person on one trial of a decision task (see Fig. 1).[2] The *x*-axis represents time. As can be seen in Fig. 1, the diffusion model distinguishes between the actual decision process (represented by *parameters v* and *a*) and a nondecision period (represented by *parameter t*0). During the actual decision process, the participant makes his or her decision on the basis of the information presented by the stimulus. The nondecision period includes processes that are applied before and after the actual decision (e.g., the perceptual encoding of the stimulus and the motor execution of a key press to indicate the decision).

The person's decision process (see the sample decision path in Fig. 1) begins at a point *z* (i.e., beginning at point *z*, the person accumulates systematic and random information

[2] It is possible to extend the basic diffusion model (including parameters *v*, *a*, *t*0, and *z*) that we describe here in order to allow for variability in trial-to-trial performance within a participant and an experiment. This extended diffusion model includes the following additional parameters: the intertrial variability of the (relative) starting point (i.e., parameter *szr*), the intertrial variability of the drift (i.e., parameter *sv*), and the intertrial variability of nondecisional components (i.e., parameter *st*0). In order to avoid unnecessarily complicating the description of the diffusion model, we explain only the basic diffusion model as it is used in our study.

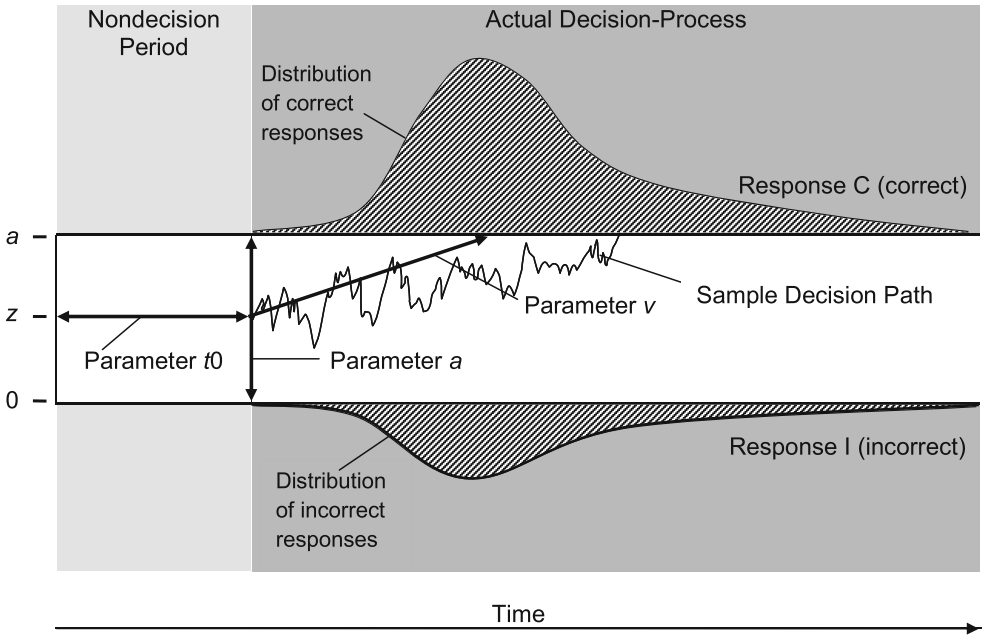

Fig. 1 Diffusion process underlying the diffusion model (cf. Schmitz & Voss, 2012). The diffusion model distinguishes between the nondecision parameter (parameter *t*0) and parameters from the actual decision process (parameters *v* and *a*). The *x*-axis represents time and is read from left to right. The *y*-axis represents the response-related decision with two response criteria, which are placed at 0 (for incorrect responses) and *a* (for correct responses). The counter begins fluctuating as a function of the information that accumulates with time at the postulated point *z*. The accumulation of information includes systematic as well as random influences. As soon as one of the two response criteria is crossed, the decision process is terminated, and the respective response is initiated. In the sample path for one trial, the participant accumulates enough information to provide the correct response. Parameter *v* is the mean amount of accumulated information for one participant across a certain number of trials (e.g., one phase in an IAT). The distributions of correct and incorrect responses are displayed outside their respective response criteria

over time). Accordingly, the sample decision path runs like an internal counter that changes over time. The counter thereby runs in a corridor between two response criteria, which are located on the response-related decision axis (i.e., the *y*-axis). The two response criteria are placed at 0 and *a*. Each criterion represents one of the two response alternatives. In our example, the upper criterion (i.e., *a*) corresponds to the correct response (i.e., Response C), whereas the lower criterion corresponds to the incorrect response (i.e., Response I). The counter is driven in opposite directions by information supporting the two different decisional outcomes (i.e., the response alternatives). As soon as one of the two response criteria (i.e., the upper criterion *a* or the lower criterion 0) is crossed, the decision process is terminated, and the response linked to the respective response criterion is initiated. In our sample, the upper criterion is crossed and thus, the correct response is initiated (see Klauer et al., 2007; Ratcliff, Gomez, et al., 2004; Ratcliff, Thapar, et al., 2004; Schmitz & Voss, 2012). Decision tasks typically include more than one trial. Hence, after the first decision has been made, a second and even some additional decisions will usually follow. The person will thus provide the response related to the first decision and will then begin again by encoding the next stimulus by accumulating information until one of the criteria is satisfied. The person will then provide the next response, and so on. Parameter *v* represents the mean amount of information accumulated by one participant across a certain number of decision processes (e.g., several trials on an IAT). The distributions of correct and incorrect responses on several hypothetical trials are displayed outside their respective response criteria in Fig. 1.

A valuable quality of the diffusion model is that it provides detailed information about the cognitive processes underlying performance on decision tasks as used in the IAT (see Voss, Nagler, & Lerche, 2013). Through the exhaustive use of available performance data (i.e., latencies of correct and incorrect responses and the accuracy of responding; see Schmitz & Voss, 2012), the diffusion model decomposes the cognitive process involved in completing the IAT into a number of meaningful components. These components are represented by several model parameters.

The parameters of the diffusion model

As can be seen in Fig. 1, the diffusion model distinguishes between several parameters. Consistent with previous research on diffusion models (Klauer et al., 2007; Ratcliff & Rouder, 1998; Schmitz & Voss, 2012; Voss et al., 2004), three parameters can be considered to be most informative: namely, two parameters from the actual decision process (parameters *v* and *a*), and one parameter from the nondecision period (parameter *t*0). The validity of these parameters has already been successfully demonstrated in various experiments (e.g., Ratcliff, Thapar, et al., 2004; Ratcliff, Thapar, & McKoon, 2001; Thapar, Ratcliff, &

McKoon, 2003; Voss et al., 2004). To make the meaning and function of these three parameters more accessible to the reader, we will exemplarily describe the decision-making process on the compatible and incompatible phases of the extraversion IAT for a fictitious participant named Peter in the following paragraphs.

Peter's task is to press the left or the right key in order to assign stimuli (e.g., the words "talkative", "me", "shy", "others") to one of four categories (extraversion-related vs. introversion-related words and self-relevant vs. non-self-relevant words). The categories are presented in either the upper left or upper right corner of the computer screen. To fulfill the task, he has to make a decision about whether the consecutively presented stimuli belong to one of the two left or two right categories. To make this decision, he samples information. Once he has sampled enough information, his decision process is terminated, and he will initiate the pressing of the respective response key. By using Peter's reaction times and errors, it is possible to estimate the following parameters.

The construct parameter v Parameter *v* refers to Peter's performance in the actual decision process. It quantifies the efficacy with which Peter accumulates response-related information. The higher Peter's value on parameter *v*, the faster Peter has reacted while simultaneously committing only a few errors—in other words, the easier it was for Peter to make his decision.

The response caution parameter a Parameter *a* refers to the amount of evidence Peter accumulates before he makes a decision in the actual decision process. It quantifies Peter's cautiousness when making his response, called *response caution*. Response caution is a non-ability-related personality characteristic that contributes to speed–accuracy settings (Schmitz & Voss, 2012). People differ in whether they prefer a conservative response mode (i.e., respond more slowly but with high accuracy) or a more liberal response mode (i.e., respond more quickly by accepting the risk of increased errors). The higher Peter's value on parameter *a*, the more information he samples before he makes his decision—in other words, the more conservative his response mode is.

The nondecision parameter t0 In contrast to the above-mentioned parameters *a* and *v*, parameter *t*0 refers *not* to the actual decision process, but to the nondecision components in Peter's reaction time (cf. Klauer et al., 2007; Schmitz & Voss, 2012). This includes, for example, the perceptual encoding of the stimuli, task preparation, task switching, and the execution of motor responses (e.g., Klauer et al., 2007; Schmitz & Voss, 2012). The last item, in particular, has been given a great deal of empirical support (e.g., Ratcliff, Thapar, & McKoon, 2006; Voss et al., 2004). Thus, the higher Peter's value on parameter *t*0, the more time he took to press a response key (e.g., due to his ability to execute this motor response).

The new IAT effects decomposed by diffusion model analyses

The parameters (i.e., *v*, *a*, and *t*0) that result from diffusion model analyses can be computed for the compatible phase as well as for the incompatible phase of the IAT. They can be used to decompose the traditional IAT effect into the following three dissociable IAT effects: IAT_v, IAT_a, and IAT_{t0} (Klauer et al., 2007). These dissociable IAT effects can be computed as the compatibility effects for the parameters *v*, *a*, and *t*0 (Klauer et al., 2007) by subtracting the estimated parameters from the compatible phase from the estimated parameters from the incompatible phase in each case (i.e., IAT_v = parameter *v* in incompatible phase – parameter *v* in compatible phase, IAT_a = parameter *a* in incompatible phase – parameter *a* in compatible phase, and IAT_{t0} = parameter *t*0 in incompatible phase – parameter *t*0 in compatible phase). In contrast to the traditional IAT effect, these IAT effects newly developed by Klauer et al. (2007) seem to be able to differentiate between different sources of variance in the IAT results, as we explain in the following paragraphs.

The construct-related IAT effect (IAT_v) In two studies, Klauer et al. (2007) demonstrated that IAT_v from a political-attitude IAT was significantly related to attitude ratings of political standpoints and, thus, was significantly associated with the construct-related variance of the IAT. In addition, construct-related variance was not mapped onto the two other new IAT effects: IAT_a and IAT_{t0} (Klauer et al., 2007), indicating that both contain some variance from the IAT effect other than the construct-related variance.

The response-caution-related IAT effect (IAT_a) As was shown by Klauer et al. (2007), IAT_a is significantly related to so-called control IATs (i.e., IATs developed to contain method-specific variance only) and, hence, is significantly associated with the method-specific variance of the IAT effect. IAT_v and IAT_{t0} were found to play only small roles in accounting for method-related variance (Klauer et al., 2007), supporting the idea that both are due to variance other than method-specific variance.

The non-decision-related IAT effect (IAT_{t0}) Last but not least, Klauer et al. (2007) demonstrated that IAT_{t0} is not significantly associated with either construct-related or method-related variance in the IAT. Thus, it appears to capture an additional, third source of variance in the IAT.

In sum, Klauer et al. (2007) successfully showed that it is possible to use diffusion model analyses to separate construct-related variance from method-specific variance. However, until now, no empirical research has investigated whether it may also be possible to use diffusion model analyses to separate construct-related variance from faking-related

variance. Before we explain how faking-related variance might influence the new IAT effects, it is important for the reader to understand how people fake the IAT.

Faking strategies on the IAT

Recent research has indicated that faking on the IAT is driven by different faking strategies (Röhner et al., 2013). To investigate faking strategies on the IAT, Röhner et al. (2013) instructed participants either to fake or not to fake their IAT scores. The faking instructions differed according to the requested faking direction (i.e., faking low scores vs. faking high scores) and according to whether the participants had or had not been previously informed about possible faking strategies (i.e., naïve vs. informed faking). Instructing participants to fake represents the most common methodology for investigating faking behavior (Smith & McDaniel, 2012), since this methodology provides valuable insight into the extent to which people can fake and into the strategies that people apply when asked to fake (see Smith & Ellingson, 2002; Smith & McDaniel, 2012).[3]

The results of Röhner et al.'s (2013) study accordingly provided important insights into the variability of IAT faking strategies. Röhner et al., (2013) found that the choice of faking strategy depended on whether participants were requested to fake *low scores* or *high scores* (i.e., the faking direction) and on participants' preexisting knowledge of faking strategies (i.e., whether participants had or had not been informed about possible faking strategies). Successful faking is driven by strategically slowing down or accelerating on either the compatible or the incompatible phase (Röhner et al., 2013; see also Fiedler & Bluemke, 2005). *Naïve fakers* (i.e., participants asked to fake without being informed about possible faking strategies) successfully fake *low scores* on the IAT by slowing down on the compatible phase and fake *high scores* by accelerating on the compatible phase (Röhner et al., 2013). *Informed fakers* (i.e., participants asked to fake after being informed about possible faking strategies) slow down on the compatible phase to fake *low scores* and slow down on the incompatible phase to fake *high scores* (Röhner et al., 2013).

The influence of faking on the new IAT effects

How might faking-related variance affect the newly developed IAT effects? This has not been investigated until now. Nevertheless, some ideas can be drawn from the literature.

[3] Note that using this methodology was appropriate for addressing questions about the extent to which people are able to fake and what strategies they use when faking. If we had been interested in people's motivation to fake in the applied settings, we would have used another methodology, because motivating participants to fake by instructing them to do so can serve only as an approximation of the motivation to fake outside the laboratory.

Faking should not affect IAT$_v$ Since the IAT$_v$ represents the IAT's construct-related variance, it should not be affected at all by faking-related variance. In line with Klauer et al.'s (2007) suggestions, faking might have an impact on parameter *a* and its compatibility effect IAT$_a$ (i.e., participants' response caution), as well as on parameter *t*0 and its compatibility effect IAT$_{t0}$ (i.e., the time needed to provide a motor response outside of the decision process). By contrast, faking should not affect IAT$_v$, because both of the parameters that are likely related to faking variance (i.e., parameters *a* and *t*0) are automatically partialed out of IAT$_v$ (see also Klauer et al., 2007). In this vein, Wagenmakers (2009) also supposed that when diffusion model analyses are used, unwanted strategic variance (e.g., faking-related variance) is filtered out of IAT$_v$.

Faking might affect IAT$_a$ Faking strategies (i.e., the acceleration or slowing down of reaction times in certain IAT phases; see Röhner et al., 2013) might be caused by the adaptation of speed–accuracy settings (see Fiedler & Bluemke, 2005). To be more specific, fakers might slow down their reaction times in an IAT phase by responding especially cautiously on it, and might accelerate their reaction times in an IAT phase by applying a more liberal response caution to it. *Naïve and informed fakers of low scores* tend to use the strategy of slowing down on the compatible phase (Röhner et al., 2013). They might slow down on the compatible phase by applying a higher response caution on that phase than on the incompatible one (see Fiedler & Bluemke, 2005; Klauer et al., 2007). This, in turn, should result in a higher parameter *a* in the compatible phase and a lower parameter *a* in the incompatible phase. As a result, the compatibility effect IAT$_a$ should decrease for participants who are asked to fake low scores. *Naïve fakers of high scores* tend to use the strategy of accelerating on the compatible phase, whereas *informed fakers of high scores* tend to use the strategy of slowing down on the incompatible phase (Röhner et al., 2013). Again, participants might do this due to variability in response caution in those phases (see Fiedler & Bluemke, 2005; Klauer et al., 2007). Both strategies (i.e., accelerating in the compatible phase and slowing down in the incompatible one) should result in a higher parameter *a* in the incompatible phase than in the compatible phase. Thus, faking high scores on the IAT should lead to an increase in the value of IAT$_a$, because the faking strategies are mirror inversions of those that result from faking low (see Röhner et al., 2013). As a consequence, faking might affect the compatibility effect IAT$_a$, because it should decrease for participants asked to fake low scores and should increase for those asked to fake high scores.

Faking might affect IAT$_{t0}$ Faking strategies might also be caused by motor-response adaptations that occur outside the actual decision process (see Klauer et al., 2007). To be more specific, fakers might slow down their reaction times in an

IAT phase by delaying their key-pressing motor response, and might accelerate their reaction times in an IAT phase by pressing the key as quickly as possible. *Naïve and informed fakers of low scores* might implement the strategy of slowing down in the compatible phase by delaying their response execution in that phase more than in the incongruent one. This should cause higher $t0$ parameters in the compatible phase and lower $t0$ parameters in the incompatible phase. As a result, the IAT_{t0} compatibility effect should decrease for participants who are asked to fake low scores. By contrast, *naïve and informed faking of high scores* on the IAT should lead to increases in the value of IAT_{t0}, because the faking strategies are mirror-inverted versions of those for faking low (see Röhner et al., 2013). Thus, faking might affect the IAT_{t0} compatibility effect, because it should decrease for participants asked to fake low scores and should increase for those asked to fake high scores.

The IAT represents a very popular measure (Bosson, Swann, & Pennebaker, 2000; Rudolph, Schröder-Abé, Schütz, Gregg, & Sedikides, 2008), and its fakeability has already been well-documented (e.g., Fiedler & Bluemke, 2005; Röhner et al., 2011, 2013; Steffens, 2004). Thus, it would be useful to be able to separate construct-related variance from faking-related variance on the IAT. However, no previous research has tested this possibility by investigating the impacts of faking on the three dissociable IAT effects. In our study, we thus tried to use diffusion model analyses to assess whether it would be possible to separate construct-related variance from variance caused by faking. Our study's hypotheses are summarized in the following points.

1. Since IAT_v is supposed to be influenced exclusively by construct-specific variance (see Klauer et al., 2007; Wagenmakers, 2009), we expected *no impact of faking instructions on* IAT_v (i.e., on the construct-specific variance of the IAT).
2. By contrast, since participants typically adapt their reaction times in order to fake the IAT (Röhner et al., 2013), and this manipulation of reaction times might be due to an adaptation of their response caution (i.e., parameter a), we expected the faking instructions to impact the compatibility effect IAT_a. IAT_a was expected to *decrease for participants asked to fake low scores*, due to higher response caution in the compatible than in the incompatible phase. By contrast, IAT_a was expected to *increase for participants asked to fake high scores* on the IAT, due to higher response caution in the incompatible than in the compatible phase.
3. Since reaction time adaptations can also be caused by slower or faster motor responses that hail from outside the decision process (i.e., parameter $t0$), we furthermore expected the faking instructions also to impact IAT_{t0}. IAT_{t0} was expected to *decrease for participants asked to fake low scores*, because they were expected to delay their motor responses more in the compatible phase than in the incompatible phase. By contrast, it was expected to *increase for participants asked to fake high scores* on the IAT, because they were expected to delay their motor responses more in the incompatible phase than in the compatible phase.

Method

Data for reanalysis

To examine the process components in the IAT under faking, we reanalyzed a published data set that was collected to investigate the behavior of fakers on the IAT (Röhner et al., 2013). We decided to reanalyze this data set for several reasons: First, this previous study had investigated the faking of high and low scores on the IAT. Since we were interested in the impact of faking low and high scores on the dissociable IAT effects, it was necessary that both possible faking directions were contained in the primary data set. Second, naïve faking and faking after being informed about faking strategies were included in the data set. Since we wanted to investigate the impact of knowledge about faking strategies on the dissociable IAT effects, we needed a primary data set in which naïve and informed faking could be compared with each other. Third, the IAT used in the previous study (i.e., the extraversion IAT) is a frequently used and very popular IAT (e.g., Grumm & von Collani, 2007; Schmukle, Back, & Egloff, 2008; Steffens & Schulze-König, 2006). Last but not least, as 84 participants were included in this study an a priori power analysis using G*Power 3.1.7 (Faul, Erdfelder, Lang, & Buchner, 2007) revealed a power of nearly 100% (.998) for ANOVAs with repeated measures to detect a moderate effect size. The minimum acceptable sample size was determined to be $N = 51$ related to a power level of .950.

As is detailed in Röhner et al. (2013), a total of 84 volunteers (64 female, 20 male; 74 students) from Chemnitz University of Technology participated in the study in exchange for personal feedback and partial course credit. Their mean age was 22.37 years ($SD = 4.45$). Participants were randomly assigned to one of three conditions: (a)a control group, (b)a faking condition LH (faking low scores first, and then faking high scores), or (c)a faking condition HL (faking high scores first, and then faking low scores). All participants completed an extraversion IAT (Back et al., 2009) a total of three times. After all participants completed the IAT once without faking instructions (i.e., baseline assessment), participants in the control group completed the IATs two more times without further instructions. Participants in the

faking conditions were asked to fake the IAT first with no information about the IAT's rationale or faking strategies (*naïve faking*). Afterwards, they were told how to fake the IAT and were asked to fake the IAT again (*informed faking*). Specifically, participants in the LH faking condition faked low scores under the naïve faking condition and high scores under the informed faking condition. Participants in the HL faking condition faked high scores under the naïve faking condition and low scores under the informed faking condition.

Analytical approach

We decided to use the command-line program *fast-dm* (Voss & Voss, 2007, 2008), which can be downloaded from the website www.psychologie.uni-heidelberg.de/ae/meth/fast-dm/. We did not use the program called DMAT (Vandekerckhove & Tuerlinckx, 2007, 2008) because DMAT, in contrast to fast-dm, requires a sufficiently high number of correct and incorrect trials (see Voss, Nagler, et al., 2013), whereas its efficiency is comparably low. The program EZ (Wagenmakers, van der Maas, Dolan, & Grasman, 2008; Wagenmakers, van der Maas, & Grasman, 2007), in contrast to fast-dm, does not estimate the relevant parameters but rather estimates only the scores that correspond to them. Thus, we used fast-dm because it met our needs best.

Pretreatment of the data set In fast-dm, the degree of correspondence between the observed cumulative distribution and the predicted cumulative distribution is quantified by the KS statistic (Kolmogorov, 1941), which is known to provide robust estimates even in the presence of minor outliers (Voss, Nagler, et al., 2013; Voss & Voss, 2008). Thus, we followed the recommendation (see Voss, Nagler, et al., 2013; Voss & Voss, 2008) to remove outliers from the individual reaction time distribution only if participants had reaction times below 200 or above 5,000 ms. This led to the exclusion of 156 trials (i.e., 0.3% of the trials).

Parameter estimation We used fast-dm to estimate the independent diffusion models for each participant (N = 84) and each combined phase type (i.e., compatible vs. incompatible phase) within every measurement occasion (i.e., baseline, retest/naïve faking, and retest/informed faking). Altogether, we thus computed 504 diffusion model analyses (i.e., 84 participants × 2 combined phase types × 3 measurement occasions). Each diffusion model analysis was based on about 96 trials (i.e., exactly 96 trials in the absence of outlier trials and an average of 92 trials when outliers were excluded).

On the basis of the respective trials, we used fast-dm to estimate the values that best fit the parameters *a*, *v*, and *t*0 to explain the observed reaction-time distributions for correct and incorrect responses. Voss et al. (2004) provide the mathematical details of parameter estimation in fast-dm in their study's Appendix.

We followed advice from existing studies on diffusion models to ensure the reliability of our model parameters. Thus, to reduce uncertainty in the parameters, only parameters *v*, *a*, and *t*0 were allowed to vary freely, whereas the starting point *z* was fixed to *a*/2 (Schmitz & Voss, 2012) and the difference in response-execution speed (i.e., parameter *d*) was fixed to zero (Voss, Voss, & Klauer, 2010). Given that the intertrial-variability parameters of the extended diffusion model (i.e., parameters *szr*, *sv*, and *st*0) are usually not very reliable (Schmitz & Voss, 2012), they were excluded as we needed only the basic diffusion model parameters for our investigation. To exclude those parameters, we set them to zero before running fast-dm.

Checking the model fit To verify whether the estimated parameters could explain the empirical data, a Kolmogorov backward equation was used to test the plausibility of the data. Here, small values of *p* indicate that the diffusion model cannot account for the data (Voss, Nagler, et al., 2013; for further details, see Voss & Voss, 2007, 2008). On the basis of Voss, Nagler, et al. (2013), we discarded a participant's data before running further analyses if they showed a poor fit to the model (i.e., p values < .05). Altogether, three participants were excluded because their data showed a poor fit. This led to a remaining sample of N = 81 participants (i.e., n = 28 in the control group, n = 25 in the LH faking condition, and n = 28 in the HL faking condition). We used fast-dm to multiply all incorrect responses by –1 and to summarize the empirical and theoretical reaction time distributions of correct and incorrect responses in one graph. Graphical displays of the model fits are presented in Fig. 2a and b.

Computation of the compatibility effects On the basis of the parameters estimated in the diffusion model analyses, we computed the compatibility effects IAT_v, IAT_a, and IAT_{t0} by subtracting the estimated parameters for the compatible phase from those for the incompatible phase (see Klauer et al., 2007). Since we computed these compatibility effects for each participant at every measurement occasion, a total of 243 compatibility effects resulted from these computations.

Analyzing the compatibility effect After fitting the data to the diffusion model, estimating the parameters, and computing the compatibility effects, we used analyses of variance (ANOVAs) with repeated measures to test our hypotheses. For all ANOVAs, we set α to .05. First, we computed 3 (measurement occasion) × 3 (experimental group) ANOVAs with repeated measures on the respective compatibility effects (i.e., IAT_v, IAT_a, and IAT_{t0}) to investigate whether faking would have an impact on those IAT effects. Second, to take a closer look at the results, we additionally computed 3 (measurement occasion) × 3 (experimental group) ANOVAs with repeated measures on each of the parameters (i.e., parameters *v*, *a*, and

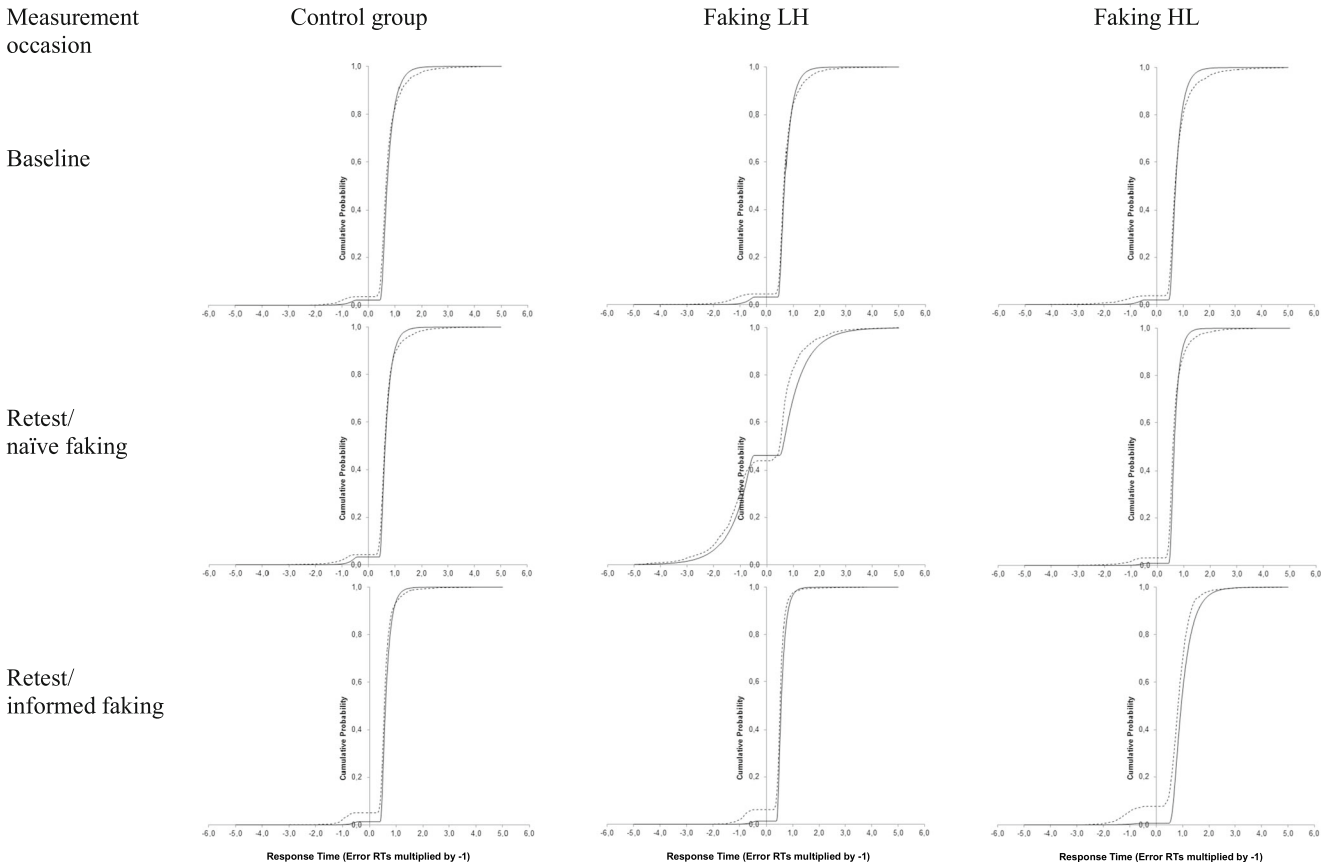

Fig. 2 **a** Overlaid predicted (parameter-based) and observed (empirical)cumulative distribution functions (cdfs) for all experimental groups and measurement occasions in the compatible IAT phase. *Continuous lines* represent the predicted cdfs and *dashed lines* represent the observed cdfs. The plotted functions are joint distributions of correct and incorrect responses. Negative values on the *x*-axis are the latencies of error responses (multiplied by –1) and are plotted on the *left side*. Positive values on the *x*-axis are the latencies of correct responses and are plotted on the *right side*. **b** Overlaid predicted (parameter-based) and observed (empirical) cumulative distribution functions (cdfs) for all experimental groups and measurement occasions in the incompatible IAT phase. *Continuous lines* represent the predicted cdfs, and *dashed lines* represent the observed cdfs. The plotted functions are joint distributions of correct and incorrect responses. Negative values on the *x*-axis are the latencies of error responses (multiplied by –1) and are plotted on the *left side*. Positive values on the *x*-axis are the latencies of correct responses and are plotted on the *right side*

*t*0), separately for the compatible and incompatible phases in each case.

Results

Effects of faking on the construct-related IAT effect (IAT_v)

To determine whether faking would affect IAT_v, we computed a 3 (measurement occasion) × 3 (experimental group) ANOVA with repeated measures on IAT_v. As expected, the main effect of group, $F(2, 78) = 1.11$, $p = .335$, $\omega^2 = .00$, was nonsignificant. Interestingly, the main effects of measurement occasion, $F(1.89, 147.67) = 3.67$, $p = .030$, $\omega^2 = .06$, and the interaction effect, $F(3.79, 147.67) = 16.91$, $p < .001$, $\omega^2 = .43$, were medium to large in size and significant.

Between- and within-group comparisons can be found in Table 1. An examination of the between- and within-group comparisons revealed that naïve faking of low and high scores as well as informed faking of high scores had influences on IAT_v. In those cases, IAT_v increased or decreased according to the requested faking direction (i.e., low or high scores, respectively).[4] Interestingly, one condition showed no (significant) influence of faking on IAT_v (i.e., informed faking of low scores). Naïve faking of low scores led to significantly higher values of IAT_v, whereas naïve faking of high scores led to significantly lower values of IAT_v. Whereas informed faking

[4] Please note that because of the computation of the compatibility effects, negative values of IAT_v indicate a strong association between the self and extraversion (as requested for the faking of high scores), and positive values of IAT_v indicate a weak association between the self and extraversion (as requested for the faking of low scores).

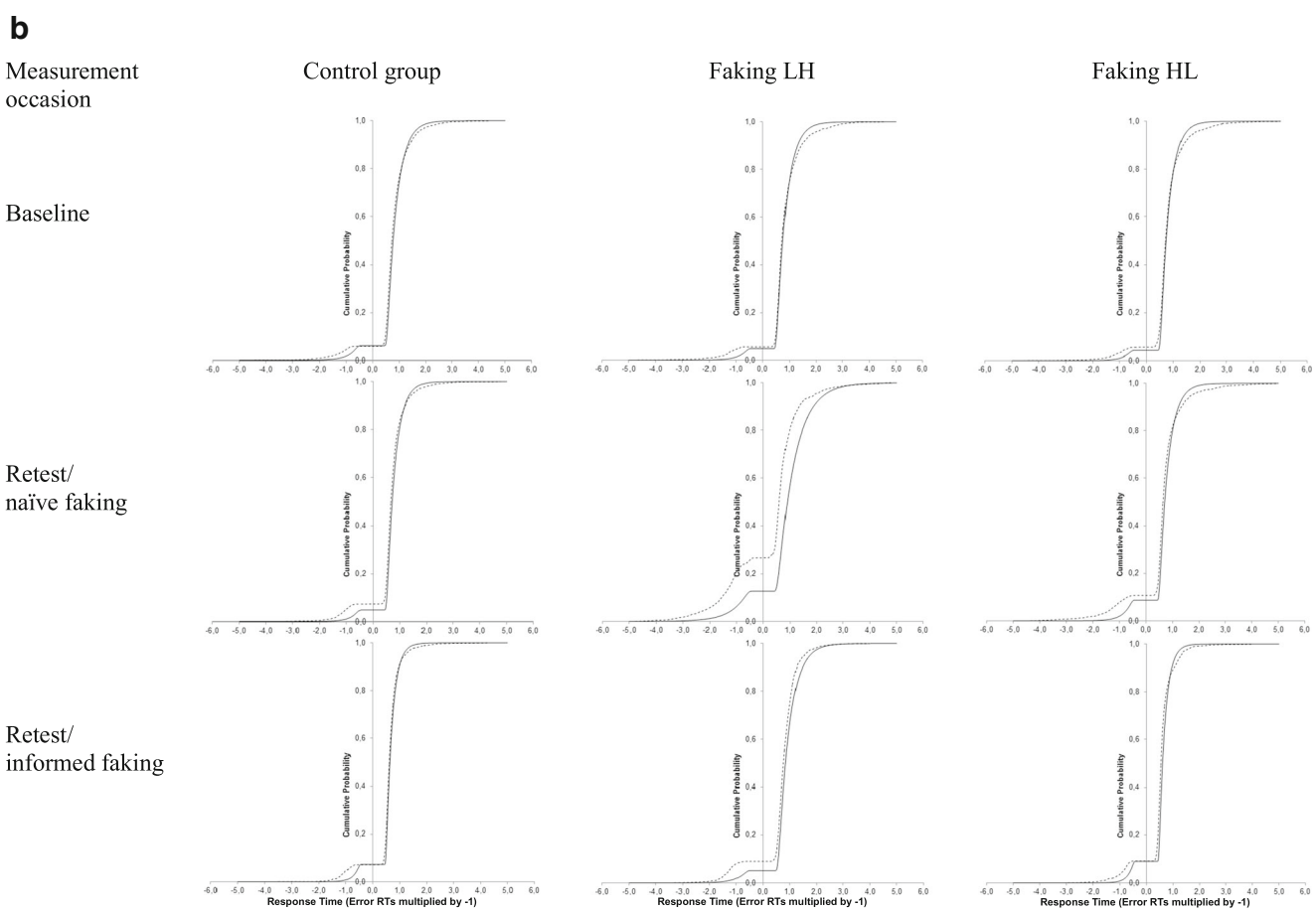

Fig. 2 (continued)

of high scores led to significantly lower values of IAT_v, informed faking of low scores did not lead to (significantly) higher values of IAT_v. Thus, the impact of faking on IAT_v depended on the specific faking condition (i.e., naïve vs. informed faking and faking low vs. high scores).

To take a closer look at the results, we additionally computed 3 (measurement occasion) × 3 (experimental group) ANOVAs with repeated measures on parameter *v* for the compatible and incompatible phases. For parameter *v* in the compatible phase, the main effects of measurement occasion,

Table 1 Post hoc comparisons of the construct-related IAT effect (IAT_v) and the construct parameter *v* from both IAT phases in the diffusion model of the extraversion IAT

Measurement Occasion	IAT_v			Parameter *v* from the Compatible IAT Phase			Parameter *v* from the Incompatible IAT Phase		
	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*
Baseline	-0.55_{a1} (0.99)	-0.41_{a1} (0.95)	-0.42_{a1} (1.04)	2.14_{a1} (0.80)	2.08_{a1} (0.88)	2.22_{a1} (1.12)	1.59_{a1} (0.54)	1.68_{a1} (0.78)	1.80_{a1} (0.65)
Retest/ Naïve faking	-0.40_{a1} (1.01)	0.87_{b2} (1.21)	-1.30_{c2} (1.46)	2.20_{a1} (0.73)	0.08_{b2} (0.95)	2.80_{a2} (1.54)	1.80_{a1} (0.81)	0.95_{b2} (1.58)	1.50_{a1} (0.94)
Retest/Informed faking	-0.75_{ab1} (1.17)	-1.30_{b3} (1.76)	-0.12_{a1} (1.35)	2.65_{a1} (1.29)	2.84_{a3} (1.70)	1.93_{a1} (1.26)	1.90_{a1} (0.81)	1.54_{a1} (0.86)	1.80_{a1} (0.89)

Faking LH represents the faking condition in which low scores were faked first and then high scores were faked; Faking HL represents the faking condition in which high scores were faked first and then low scores were faked; $N = 81$ (i.e., $n = 28$ in the control group, $n = 25$ in the LH faking condition, and $n = 28$ in the HL faking condition); different alphabetic subscripts indicate significant differences between experimental groups (i.e., columns); different numeric subscripts identify significant differences between measurement occasions (i.e., rows) at $p < .05$

$F(1.87, 145.57) = 13.11$, $p < .001$, $\omega^2 = .22$, and group, $F(2, 78) = 4.90$, $p = .010$, $\omega^2 = .09$, and the interaction effect, $F(3.73, 145.57) = 25.44$, $p < .001$, $\omega^2 = .53$, were medium to large in size and significant. For parameter *v* in the incompatible phase, the main effect of measurement occasion, $F(1.92, 149.34) = 4.68$, $p = .012$, $\omega^2 = .08$, and the interaction effect, $F(3.83, 149.34) = 2.67$, $p = .036$, $\omega^2 = .07$, were medium in size and significant, too. The main effect of group, $F(2, 78) = 2.25$, $p = .112$, $\omega^2 = .03$, remained nonsignificant. These results highlight the idea that faking affects parameter *v* in the compatible as well as the incompatible phase. However, the effect of faking on parameter *v* is more pronounced in the compatible phase.

Examining the between- and within-group comparisons for parameter *v* provided additional information about the influence of faking on parameter *v* (see Table 1). In the compatible phase, faking low led to a significantly lower parameter *v* under naïve faking, but not under informed faking. Faking high led to a significantly higher parameter *v* under both naïve and informed faking (see Table 1). In the incompatible phase, faking low led to a significant decrease in parameter *v* under naïve but not under informed faking. Both naïve and informed faking of high scores led to a nonsignificant decrease in parameter *v*.

Thus, IAT_v could not be considered to be an exclusive indicator of construct-related variance, because faking intentions for the most part had an impact on parameter *v*, and therefore also on IAT_v.

Effects of faking on the response-caution-related IAT effect (IAT_a)

To determine whether faking would affect IAT_a, we computed a 3 (measurement occasion) × 3 (experimental group) ANOVA with repeated measures on IAT_a. As expected, the main effect of measurement occasion, $F(1.67, 129.96) = 4.82$, $p = .014$, $\omega^2 = .07$, the main effect of group, $F(2, 78) = 7.84$, $p = .001$, $\omega^2 = .14$, and the interaction, $F(3.33, 129.96) = 4.50$, $p = .004$, $\omega^2 = .13$, were medium to large in size and significant.

The between- and within-group comparisons (see Table 2) showed that, surprisingly, IAT_a was affected by only one faking condition. As expected, IAT_a significantly decreased when informed participants were asked to fake low scores. Surprisingly, the informed faking of high scores, as well as the naïve faking of low and high scores, did not lead to significant differences in IAT_a.

To take a closer look at the results, we additionally computed 3 (measurement occasion) × 3 (experimental group) ANOVAs with repeated measures on parameter *a* for the compatible and incompatible phases. For parameter *a* in the compatible phase, the main effect of measurement occasion, $F(1.48, 115.13) = 1.30$, $p = .271$, $\omega^2 = .01$, and the main effect of group, $F(2, 78) = 2.94$, $p = .059$, $\omega^2 = .04$, remained nonsignificant. However, the interaction effect, $F(2.95, 115.13) = 4.05$, $p = .009$, $\omega^2 = .10$, was medium in size and significant. For parameter *a* in the incompatible phase, the main effect of measurement occasion, $F(1.89, 147.50) = 4.90$, $p = .010$, $\omega^2 = .08$, and the main effect of group, $F(2, 78) = 4.73$, $p = .012$, $\omega^2 = .08$, were medium in size and significant. The interaction effect, $F(3.78, 147.50) = 2.05$, $p = .094$, $\omega^2 = .05$, remained nonsignificant.

The results of the between- and within-group comparisons for parameter *a* document the influence of the informed faking of low scores on IAT_a. As expected, the informed faking of low scores led to a significant increase in parameter *a* in the compatible phase and to a decrease in the incompatible phase (see Table 2). Surprisingly, the informed faking of high scores as well as the naïve faking of low and high scores did not affect parameter *a* in either IAT phase.

Thus, IAT_a appeared to capture some faking-related variance. Surprisingly, it did so only when participants had been informed about faking strategies on the IAT and were asked to fake low scores. However, the informed faking of high scores as well as the naïve faking of low or high scores did not affect IAT_a.

Effects of faking on the non-decision-related IAT effect (IAT_{t0})

To determine whether faking would affect IAT_{t0}, we computed a 3 (measurement occasion) × 3 (experimental group) ANOVA with repeated measures on IAT_{t0}. The main effect of measurement occasion, $F(1.85, 143.98) = 4.10$, $p = .021$, $\omega^2 = .07$, was medium in size and significant. Interestingly, the main effect of group, $F(2, 78) = 1.41$, $p = .251$, $\omega^2 = .01$, and the interaction, $F(3.69, 143.98) = 2.06$, $p = .095$, $\omega^2 = .04$, remained nonsignificant.

The results for the between- and within-group comparisons (see Table 3) showed that, unexpectedly, there were no significant differences between and within groups on IAT_{t0} in any of the faking conditions (i.e., naïve vs. informed faking and faking high vs. low scores).

To take a closer look at the results, we additionally computed 3 (measurement occasion) × 3 (experimental group) ANOVAs with repeated measures on parameter *t*0 for the compatible and incompatible phases. For parameter *t*0 in the compatible phase, the main effects of measurement occasion, $F(1.75, 136.25) = 2.70$, $p = .078$, $\omega^2 = .04$, and group, $F(2, 78) = 0.51$, $p = .602$, $\omega^2 = .00$, and their interaction, $F(3.49, 136.25) = 0.83$, $p = .493$, $\omega^2 = .00$, remained nonsignificant. For parameter *t*0 in the incompatible phase, the main effect of measurement occasion, $F(1.85, 143.93) = 1.49$, $p = .230$, $\omega^2 = .01$, the main effect of group, $F(2, 78) = 0.27$, $p = .761$, $\omega^2 = .00$, and the interaction, $F(3.69, 143.93) = 0.84$, $p = .497$, $\omega^2 = .01$, also remained nonsignificant.

The between- and within-group comparisons for parameter *t*0 illustrated that there was no influence of faking on IAT_{t0}. In

Table 2 Post hoc comparisons of the response-caution-related IAT effect (IAT_a) and the response-caution parameter *a* from both IAT phases in the diffusion model of the extraversion IAT

	IAT_a			Parameter *a* from the Compatible IAT Phase			Parameter *a* from the Incompatible IAT Phase		
Measurement Occasion	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*
Baseline	-0.06_{a1} (1.12)	0.13_{a1} (0.62)	-0.01_{a1} (0.47)	1.78_{a1} (0.92)	1.64_{a1} (0.51)	1.75_{a1} (0.46)	1.72_{a1} (0.72)	1.77_{a1} (0.50)	1.74_{a1} (0.49)
Retest/ Naïve faking	0.13_{a1} (0.91)	0.17_{a1} (0.92)	-0.14_{a1} (1.01)	1.54_{a1} (0.59)	1.86_{a1} (0.48)	1.71_{a1} (0.88)	1.66_{a1} (0.84)	2.03_{a1} (0.90)	1.57_{a12} (0.48)
Retest/Informed faking	-0.29_{a1} (0.72)	0.37_{a1} (1.16)	-1.33_{b2} (2.06)	1.63_{a1} (0.74)	1.52_{a1} (1.10)	2.60_{b2} (2.07)	1.34_{a1} (0.32)	1.89_{b1} (0.86)	1.28_{a2} (0.69)

Faking LH represents the faking condition in which low scores were faked first and then high scores were faked; Faking HL represents the faking condition in which high scores were faked first and then low scores were faked; $N = 81$ (i.e., $n = 28$ in the control group, $n = 25$ in the LH faking condition, and $n = 28$ in the HL faking condition); different alphabetic subscripts indicate significant differences between experimental groups (i.e., columns); different numeric subscripts identify significant differences between measurement occasions (i.e., rows) at $p < .05$

other words, parameter *t*0 was not influenced by naïve faking, informed faking, the faking of high scores, or the faking of low scores in the compatible or the incompatible phase (see Table 3). Thus, interestingly, IAT_{t0} did not appear to capture any faking-related variance at all.[5]

Discussion

In the present study, we investigated whether diffusion model analyses could be used to separate construct-related variance from faking-related variance on the IAT. The results showed the advantage of using diffusion model analyses to compute the dissociable IAT effects IAT_v, IAT_a, and IAT_{t0}, since they allowed important insights into the different sources of variance that are related to the IAT effect.

Does faking have an impact on the construct-related IAT effect (IAT_v)?

Interestingly, our results showed that faking instructions had an impact on IAT_v. However, faking did not affect IAT_v in the same way in every faking condition. Faking had an impact on IAT_v when participants were asked to fake low scores or high scores without being informed about possible faking strategies, or when they had been informed about possible faking strategies and had to fake high scores. In those cases, faking low led to an increase and faking high to a decrease in IAT_v. Only one condition showed no (significant) influence of faking on IAT_v (i.e., informed faking of low scores).

Parameter *v* in the compatible and incompatible phases provided additional information about the results. First, the impact of faking was more pronounced for parameter *v* in the compatible than in the incompatible phase. This finding is in line with recent research that has shown that in most cases, people who are able to fake successfully use strategies to manipulate their reaction times in the compatible phase (see Röhner et al., 2013). Second, the naïve faking of low scores was associated with decreases in parameter *v* in the compatible as well as in the incompatible IAT phase. The latter might sound counterintuitive, since it does not mirror a successful faking strategy in this condition (Röhner et al., 2013). However, it might be explained by Röhner et al.'s (2013) results, which showed that naïve fakers of low scores also (unsuccessfully) tried to fake low scores by committing errors in an unsystematic manner (i.e., on both IAT phases). Taken together, these results indicate that parameter *v* and IAT_v seem to capture some of the variance caused by faking strategies.

These results on the impact of faking on IAT_v might at first glance appear somewhat disappointing, since the construct-related IAT effect was shown to be affected by faking. However, they provide important insights into the faking process itself and into the faking strategies that are used on the IAT. The results show that, in addition to the preexisting implicit associations that are supposed to be measured with the IAT (see Greenwald et al., 1998), faking intentions also

[5] It seems plausible that faking might be indicated by a pronounced misfit of the diffusion model (see Klauer et al., 2007). Thus, we tested whether faking would decrease the *p* value from the Kolmogorov backward equation. Only the naïve faking of low scores led to somewhat smaller *p*-values. This finding might be explained by a result from recent research indicating that naïve fakers of low scores try to fake by making mistakes (Röhner et al., 2013). Increases in errors might have an impact on the plausibility of the diffusion model, and thus might really indicate faking in this case. Faking high scores and informed faking did not affect the *p* values of the model fit. Again, this might be explained by Röhner et al.'s (2013) finding that in these conditions, fakers did not try to fake by making mistakes. Thus, the model fit might provide some indication of faking when uninformed participants are asked to fake low. Detailed results can be obtained from the corresponding author upon request.

Table 3 Post hoc comparisons of the non-decision-related IAT effect (IAT_{t0}) and the nondecision parameter *t*0 from both IAT phases in the diffusion model of the extraversion IAT

	IAT_{t0}			Parameter *t*0 from the Compatible IAT Phase			Parameter *t*0 from the Incompatible IAT Phase		
Measurement Occasion	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*	Control Group *M (SD)*	Faking LH *M (SD)*	Faking HL *M (SD)*
Baseline	0.04_{a1} (0.12)	0.00_{a12} (0.07)	0.00_{a1} (0.09)	0.39_{a1} (0.09)	0.40_{a1} (0.05)	0.41_{a1} (0.09)	0.42_{a1} (0.09)	0.40_{a1} (0.07)	0.40_{a1} (0.08)
Retest/ Naïve faking	0.02_{a1} (0.10)	-0.05_{a1} (0.19)	0.02_{a1} (0.10)	0.38_{a1} (0.07)	0.42_{a1} (0.14)	0.39_{a1} (0.09)	0.40_{a1} (0.09)	0.37_{a1} (0.17)	0.40_{a1} (0.06)
Retest/Informed faking	0.06_{a1} (0.06)	0.07_{a2} (0.13)	0.01_{a1} (0.16)	0.36_{a1} (0.08)	0.35_{a1} (0.08)	0.39_{a1} (0.19)	0.41_{a1} (0.07)	0.43_{a1} (0.14)	0.40_{a1} (0.14)

Faking LH represents the faking condition in which low scores were faked first and then high scores were faked; Faking HL represents the faking condition in which high scores were faked first and then low scores were faked; $N = 81$ (i.e., $n = 28$ in the control group, $n = 25$ in the LH faking condition, and $n = 28$ in the HL faking condition); different alphabetic subscripts indicate significant differences between experimental groups (i.e., columns); different numeric subscripts identify significant differences between measurement occasions (i.e., rows) at $p < .05$

influence the ease of decision-making on the IAT (i.e., captured by IAT_v). How might this influence of faking on IAT_v be explained? Faking might to some extent result from temporary changes in people's accessible mental associations. For example, people could try to take on the role of an extraverted person by telling themselves, "For the next few minutes, I will be an extraverted person", or the role of an introverted person by telling themselves, "For the next few minutes, I will be an introverted person". It has already been shown that it is possible to create temporary mental associations (De Houwer et al., 2007). Such temporary mental associations may be helpful for faking, especially under difficult faking conditions (e.g., faking without preexisting knowledge about faking strategies). A temporary change in mental associations as a result of faking might in turn influence the ease of decision-making on the IAT, and may consequently also influence IAT_v. Thus, under faking, IAT_v not only contains construct-related variance, but also faking-related variance, since the way that people's faking intentions contribute to the ease of decision-making may be similar to the way that preexisting implicit associations do.

However, it is interesting that IAT_v was affected by faking in only three out of the four different faking conditions. The informed faking of low scores had no (significant) impact on IAT_v. How might this result be explained? Several studies have revealed that faking with a recommended strategy is easier than faking without a recommended strategy (e.g., Fiedler & Bluemke, 2005; Kim, 2003; Röhner et al., 2011) and that faking low scores is easier than faking high scores (Röhner et al., 2011). Thus, faking low scores with a recommended strategy for how to do so represents the easiest faking condition. Most likely, participants in this (easy) faking condition did not need to establish a temporary mental association to fake successfully, and hence, faking had no (significant) impact on IAT_v in this faking condition.

Does faking have an impact on the response-caution-related IAT effect (IAT_a)?

The impact of faking on IAT_a was a mirror inversion of the impact of faking on IAT_v. IAT_a was not affected by faking when participants faked low scores or high scores without being informed about possible faking strategies, or when they were informed and were asked to fake high scores. In other words, IAT_a was affected only when participants faked low scores after being informed about faking strategies (i.e., the easiest faking condition). In this condition, faking low scores was reflected by the expected decrease in the value of IAT_a. As a consequence, IAT_a may indeed capture some of the faking-related variance, but only for informed fakers of low scores (i.e., not for informed fakers of high scores or naïve fakers of low and high scores).

The results for parameter *a* in the compatible and incompatible phases might provide additional information about this result. IAT_a was affected by the informed faking of low scores (i.e., the easiest faking condition) because of a more conservative speed–accuracy setting in the compatible phase and a more liberal one in the incompatible phase. By contrast, parameter *a* in the compatible and incompatible phases was not significantly influenced by faking in the more difficult faking conditions (i.e., faking without being informed about possible faking strategies or faking high scores).

How can this be explained? Again, this result sheds more light on the faking process itself. The finding reveals that when participants were confronted with a comparably easy faking condition (i.e., the informed faking of low scores), they did not have to temporarily change their mental associations to fake successfully (as in the more difficult faking conditions, such as naïve faking or faking high scores). Instead, they were able to simply manipulate their response caution in order to fake successfully (i.e., to apply higher response caution in the compatible than in the incompatible phase).

Does faking have an impact on the non-decision-related IAT effect (IAT_{t0})?

Surprisingly, IAT_{t0} was not affected at all by faking in any faking condition (i.e., under naïve or informed faking of low or high scores). The results for parameter *t*0 in the compatible and incompatible phases demonstrated that IAT_{t0} values (i.e., the response time outside of the actual decision process) were nearly equal across all experimental conditions. Thus, IAT_{t0} does not capture any faking-related variance at all.

Remembering that IAT_{t0} is assumed to reflect components that lie outside of the actual decision process, this result again provides important insights into the faking process. The result indicates that fakers do not manipulate their reaction times outside of the decision process, but instead manipulate the decision process itself by temporarily changing their accessible mental associations (thus affecting IAT_v) or, under comparably easy faking conditions, by manipulating their response caution (thus affecting IAT_a). This finding in turn might to some extent explain why it is so difficult to detect fakers (see Fiedler & Bluemke, 2005; Röhner et al., 2013). On the one hand, we already know that faking indices have a high risk of misclassification, since they are based solely on detecting some kind of deceleration in a person's response rate (i.e., a strategy not applied by all fakers; see Röhner et al., 2013). On the other hand, as was shown in this study, fakers do not manipulate their reaction times outside of the actual decision process (e.g., by delaying their motor execution of the key pressing), but instead manipulate the decision process itself (e.g., by manipulating the ease of decision-making by establishing helpful mental associations). Thus, the risk of misclassification associated with existing faking indices (see Röhner et al., 2013) might additionally be explained by the possibility that faking behavior and nonfaking behavior may simply not be distinctive enough to be clearly separated from each other by faking indices.

The assets and drawbacks of using diffusion model analyses

From an applied point of view, researchers might be somewhat deterred from using diffusion model analyses to compute these *newly developed IAT effects*. In fact, as compared with the computation of the *traditional IAT effect*, this new procedure is more complex and time-consuming, which might be seen as two relevant drawbacks. In addition, our results indicate that a supposed relevant advantage of this new method does not hold true, because faking surprisingly affected the construct-related IAT effect IAT_v. Thus, as is true for the traditional IAT effect, IAT_v also has to be interpreted with caution, since it can be confounded with faking-related variance.

Still, diffusion model analyses are also associated with at least two relevant assets. One first, great benefit of using diffusion model analyses to compute model parameters and compatibility effects is that they can be applied in order to screen data to identify some first indications of faking. This is especially important because a multitude of research has shown that the IAT can be faked (e.g., De Houwer et al., 2007; Fiedler & Bluemke, 2005; McDaniel et al., 2009; Röhner et al., 2011; Steffens, 2004). Recent research has additionally indicated that identifying IAT fakers is much more difficult than had previously been thought (Fiedler & Bluemke, 2005; Röhner et al., 2013). The naïve faking of low scores might stand out because of decreases in the *p* values associated with parameter estimation. The informed faking of low scores might be indicated by effects on IAT_a, but the naïve and informed faking of high scores might not be indicated, because faking high scores only affects the construct-related IAT effect (i.e., IAT_v). From a practical standpoint, faking low is indeed the more likely faking condition, because low scores often represent the associations that are not socially stigmatized in particularly sensitive areas, such as pedophilia, racism, stereotypes, or sexism (e.g., Agerström & Rooth, 2011; Banse, Schmidt, & Clarbour, 2010; Banse et al., 2001; Carlsson & Björklund, 2010; Gray, Brown, MacCulloch, Smith, & Snowden, 2005; Greenwald & Banaji, 1995; Greenwald et al., 1998; Latu et al., 2011).

A second important asset of diffusion model analyses, as revealed by our results, is that they provide detailed information about the cognitive processes underlying the faking process on the IAT. Although some research has tried to uncover what people actually do to fake the IAT (e.g., Agosta, Ghirardi, Zogmaister, Castiello, & Sartori, 2011; Cvencek, Greenwald, Brown, Gray, & Snowden, 2010; Röhner et al., 2013), the processes behind faking on the IAT still represent a sort of black box. The diffusion model has allowed us to peer into this black box to reveal at least three important insights into the faking process. First, people are able to change their mental associations in order to fake successfully, at least to some extent. They do so especially in difficult faking conditions (i.e., faking high scores or naïve faking conditions). Thus, in difficult faking conditions, establishing appropriate mental associations may help people to fake successfully. Second, in comparably easy faking conditions (i.e., faking low scores), fakers do not need to establish such helpful mental associations, but instead they may simply respond more cautiously or liberally in order to fake successfully. Third, faking represents a process that takes place within the decision process and not outside of it—an important finding that in turn helps to explain why detecting fakers is so difficult (see Fiedler & Bluemke, 2005; Röhner et al., 2013).

Limitations

Of course, this study is limited by the fact that we used only an extraversion IAT. The results should be replicated using IATs that measure other constructs. In addition, the participants in

this reanalyzed data set were students with an average age of 22.37 years. However, age is related to reaction times and errors (e.g., Endrass, Schreiber, & Kathmann, 2012). Since both are used for parameter estimation in diffusion model analyses, future research should replicate and extend the findings of our study by using a sample that includes older participants. Last but not least, we used only diffusion model analyses (Klauer et al., 2007). Of course, other models, such as the quad model (Conrey, Sherman, Gawronski, Hugenberg, & Groom, 2005) or the discrimination–association model (Stefanutti, Robusto, Vianello, & Anselmi, 2013), have been proposed too for decomposing the processes underlying the IAT effect. However, only within the diffusion model analyses (Klauer et al., 2007) the impact of faking on the model parameters was considered.

Summary and conclusion

At present, it is not possible to clearly separate construct- and faking-related variance on the IAT by using diffusion model analyses. As is also true for the traditional IAT effect, the construct-related IAT effect IAT_v is influenced by faking. Nevertheless, diffusion model analyses can help us understand and interpret IAT scores. First, the use of diffusion models enabled us to determine why it is not yet possible to separate construct-related variance from faking-related variance on the IAT. We showed that the ease of decision making (as captured by the construct-related IAT effect IAT_v) was systematically manipulated by fakers. To prevent faking from having an impact on the construct-related IAT effect, future research should investigate how IAT_v can be assessed more purely. Second, by using diffusion model analyses it might be possible to screen data to obtain some indication of faked IAT scores. Whereas fakers of high scores did not stand out, poorly fitting models might indicate that participants have tried to fake low scores in a naïve manner, and fakers of low scores who have knowledge about possible faking strategies might stand out because of their values on IAT_a. Last but not least, our results clearly showed that the diffusion model represents a useful tool for investigating the faking process and for expanding our understanding of how people fake the IAT.

References

Agerström, J., & Rooth, D.-O. (2011). The role of automatic obesity stereotypes in real hiring discrimination. *Journal of Applied Psychology, 96,* 790–805. doi:10.1037/a0021594

Agosta, S., Ghirardi, V., Zogmaister, C., Castiello, U., & Sartori, G. (2011). Detecting fakers of the autobiographical IAT. *Applied Cognitive Psychology, 25,* 299–306. doi:10.1001/acp.1691

Back, M. D., Schmukle, S. C., & Egloff, B. (2005). Measuring task-switching ability in the Implicit Association Test. *Experimental Psychology, 52,* 167–179. doi:10.1027/1618-3169.52.3.167

Back, M. D., Schmukle, S. C., & Egloff, B. (2009). Predicting actual behavior from the explicit and implicit self-concept of personality. *Journal of Personality and Social Psychology, 97,* 533–548. doi:10.1037/a0016229

Banse, R., Schmidt, A. F., & Clarbour, J. (2010). Indirect measures of sexual interest in child sex offenders: A multimethod approach. *Criminal Justice and Behavior, 37,* 319–335. doi:10.1177/0093854809357598

Banse, R., Seise, J., & Zerbes, N. (2001). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Zeitschrift für Experimentelle Psychologie, 48,* 145–160. doi:10.1026/0949-3946.48.2.145

Bar-Anan, Y., & Nosek, B. A. (2014). A comparative investigation of seven indirect attitude measures. *Behavior Research Methods, 46,* 668–688. doi:10.3758/s13428-013-0410-6

Bosson, J. K., Swann, W. B., Jr., & Pennebaker, J. W. (2000). Stalking the perfect measure of implicit self-esteem: The blind men and the elephant revisited? *Journal of Personality and Social Psychology, 79,* 631–643. doi:10.1037/0022-3514.79.4.631

Carlsson, R., & Björklund, F. (2010). Implicit stereotype content: Mixed stereotypes can be measured with the Implicit Association Test. *Social Psychology, 41,* 213–222. doi:10.1027/1864-9335/a000029

Conrey, F. R., Sherman, J. W., Gawronski, B., Hugenberg, K., & Groom, C. J. (2005). Separating multiple processes in implicit social cognition: The quad model of implicit task performance. *Journal of Personality and Social Psychology, 89,* 469–487. doi:10.1037/0022-3514.89.4.469

Cvencek, D., Greenwald, A. G., Brown, A. S., Gray, N. S., & Snowden, R. J. (2010). Faking of the Implicit Association Test is statistically detectable and partly correctable. *Basic and Applied Social Psychology, 32,* 302–314. doi:10.1080/01973533.2010.519236

De Houwer, J., Beckers, T., & Moors, A. (2007). Novel attitudes can be faked on the Implicit Association Test. *Journal of Experimental Social Psychology, 43,* 972–978. doi:10.1016/j.jesp.2006.10.007

Endrass, T., Schreiber, M., & Kathmann, N. (2012). Speeding up older adults: Age-effects on error processing in speed and accuracy conditions. *Biological Psychology, 89,* 426–432. doi:10.1016/j.biopsycho.2011.12.005

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G* Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39,* 175–191. doi:10.3758/BF03193146

Fiedler, K., & Bluemke, M. (2005). Faking the IAT: Aided and unaided response control on the implicit association tests. *Basic and Applied Social Psychology, 27,* 307–316. doi:10.1207/s15324834basp2704_3

Gawronski, B. (2002). What does the Implicit Association Test measure? A test of the convergent and discriminant validity of prejudice-related IATs. *Experimental Psychology, 49,* 171–180. doi:10.1026//1618-3169.49.3.171

Gray, N. S., Brown, A. S., MacCulloch, M. J., Smith, J., & Snowden, R. J. (2005). An implicit test of the associations between children and sex in pedophiles. *Journal of Abnormal Psychology, 114,* 304–308. doi:10.1037/0021-843X.114.2.304

Greenwald, A. G., & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review, 102,* 4–27. doi:10.1037/0033-295X.102.1.4

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology, 74,* 1464–1480. doi:10.1037/0022-3514.74.6.1464

Greenwald, A. G., Nosek, B. A., & Banaji, M. R. (2003a). Understanding and using the Implicit Association Test: I. An improved scoring

algorithm. *Journal of Personality and Social Psychology, 85,* 197–216. doi:10.1037/0022-3514.85.2.197

Greenwald, A.G., Nosek, B.A., & Banaji, M.R. (2003b). "Understanding and using the Implicit Association Test: I. An improved scoring algorithm": Correction to Greenwald et al. (2003). *Journal of Personality and Social Psychology*, *85*, 481. doi:10.1037/h0087889

Grumm, M., & von Collani, G. (2007). Measuring Big-Five personality dimensions with the implicit association test: Implicit personality traits or self-esteem? *Personality and Individual Differences, 43,* 2205–2217. doi:10.1016/j.paid.2007.06.032

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H., & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin, 31,* 1369–1385. doi:10.1177/0146167205275613

Kim, D.-Y. (2003). Voluntary controllability of the Implicit Association Test (IAT). *Social Psychology Quarterly, 66,* 83–96. doi:10.2307/3090143

Klauer, K. C., Voss, A., Schmitz, F., & Teige-Mocigemba, S. (2007). Process components of the Implicit Association Test: A diffusion-model analysis. *Journal of Personality and Social Psychology, 93,* 353–368. doi:10.1037/0022-3514.93.3.353

Kolmogorov, A. (1941). Confidence limits for an unknown distribution function. *Annals of Mathematical Statistics, 12,* 461–463. doi:10.1214/aoms/1177731684

Latu, I. M., Stewart, T. L., Myers, A. C., Lisco, C. G., Estes, S. B., & Donahue, D. K. (2011). What we "say" and what we "think" about female managers: Explicit versus implicit associations of women with success. *Psychology of Women Quarterly, 35,* 252–266. doi:10.1177/0361684310383811

McDaniel, M. J., Beier, M. E., Perkins, A. W., Goggin, S., & Frankel, B. (2009). An assessment of the fakeability of self-report and implicit personality measures. *Journal of Research in Personality, 43,* 682–685. doi:10.1016/j.jrp.2009.01.011

McFarland, S., & Crouch, Z. (2002). A cognitive skill confound on the Implicit Association Test. *Social Cognition, 20,* 483–510. doi:10.1521/soco.20.6.483.22977

Mierke, J., & Klauer, K. C. (2003). Method-specific variance in the Implicit Association Test. *Journal of Personality and Social Psychology, 85,* 1180–1192. doi:10.1037/0022-3514.85.6.1180

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review, 85,* 59–108. doi:10.1037/0033-295X.85.2.59

Ratcliff, R. (2014). Measuring psychometric functions with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance, 40,* 870–888. doi:10.1037/a0034954

Ratcliff, R., Gomez, P., & McKoon, G. (2004). A diffusion model account of the lexical decision task. *Psychological Review, 111,* 159–182. doi:10.1037/0033-295X.111.1.159

Ratcliff, R., & Rouder, J. N. (1998). Modeling response times for two-choice decisions. *Psychological Science, 9,* 347–356. doi:10.1111/1467-9280.00067

Ratcliff, R., & Rouder, J. N. (2000). A diffusion model account of masking in two-choice letter identification. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 127–140. doi:10.1037/0096-1523.26.1.127

Ratcliff, R., Thapar, A., Gomez, P., & McKoon, G. (2004). A diffusion model analysis of the effects of aging in the lexical-decision task. *Psychology and Aging, 19,* 278–289. doi:10.1037/0882-7974.19.2.278

Ratcliff, R., Thapar, A., & McKoon, G. (2001). The effects of aging on reaction time in a signal detection task. *Psychology and Aging, 16,* 323–341. doi:10.1037/0882-7974.16.2.323

Ratcliff, R., Thapar, A., & McKoon, G. (2006). Aging and individual differences in rapid two-choice decisions. *Psychonomic Bulletin & Review, 13,* 626–635. doi:10.3758/BF03193973

Röhner, J., Schröder-Abé, M., & Schütz, A. (2011). Exaggeration is harder than understatement, but practice makes perfect! Faking success in the IAT. *Experimental Psychology, 58,* 464–472. doi:10.1027/1618-3169/a000114

Röhner, J., Schröder-Abé, M., & Schütz, A. (2013). What do fakers actually do to fake the IAT? An investigation of faking strategies under different faking conditions. *Journal of Research in Personality, 47,* 330–338. doi:10.1016/j.jrp.2013.02.009

Rudolph, A., Schröder-Abé, M., Schütz, A., Gregg, A. P., & Sedikides, C. (2008). Through a glass, less darkly? Reassessing convergent and discriminant validity in measures of implicit self-esteem. *European Journal of Psychological Assessment, 24,* 273–281. doi:10.1027/1015-5759.24.4.273

Schmitz, F., & Voss, A. (2012). Decomposing task-switching costs with the diffusion model. *Journal of Experimental Psychology: Human Perception and Performance, 38,* 222–250. doi:10.1037/a0026003

Schmukle, S. C., Back, M. D., & Egloff, B. (2008). Validity of the five-factor model for the implicit self-concept of personality. *European Journal of Psychological Assessment, 24,* 263–272. doi:10.1027/1015-5759.24.4.263

Smith, D. B., & Ellingson, J. E. (2002). Substance versus style: A new look at social desirability in motivating contexts. *Journal of Applied Psychology, 87,* 211–219. doi:10.1037/0021-9010.87.2.211

Smith, D. B., & McDaniel, M. (2012). Questioning old assumptions: Faking and the personality–performance relationship. In M. Ziegler, C. MacCann, & R. D. Roberts (Eds.), *New perspectives on faking in personality assessment* (pp. 53–69). Oxford: Oxford University Press.

Spaniol, J., Madden, D. J., & Voss, A. (2006). A diffusion model analysis of adult age differences in episodic and semantic long-term memory retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 32,* 101–117. doi:10.1037/0278-7393.32.1.101

Stefanutti, L., Robusto, E., Vianello, M., & Anselmi, P. (2013). A discrimination–association model for decomposing component processes of the Implicit Association Test. *Behavior Research Methods, 45,* 393–404. doi:10.3758/s13428-012-0272-3

Steffens, M. (2004). Is the Implicit Association Test immune to faking? *Experimental Psychology, 51,* 165–179. doi:10.1027/1618-3169.51.3.165

Steffens, M. C., & Schulze-König, S. (2006). Predicting spontaneous Big-Five behavior with implicit association tests. *European Journal of Psychological Assessment, 22,* 13–20. doi:10.1027/1015-5759.22.1.13

Thapar, A., Ratcliff, R., & McKoon, G. (2003). A diffusion model analysis of the effects of aging on letter discrimination. *Psychology and Aging, 18,* 415–429. doi:10.1037/0882-7974.18.3.415

van Ravenzwaaij, D., van der Maas, H. L. J., Wagenmakers, E.-J. (2011). Does the name-race Implicit Association Test measure racial prejudice? *Experimental Psychology, 58*, 271–277. doi:10.1027/1618-3169/a000093

Vandekerckhove, J., & Tuerlinckx, F. (2007). Fitting the Ratcliff diffusion model to experimental data. *Psychonomic Bulletin & Review, 14,* 1011–1026. doi:10.3758/BF03193087

Vandekerckhove, J., & Tuerlinckx, F. (2008). Diffusion model analysis with MATLAB: A DMAT primer. *Behavior Research Methods, 40,* 61–72. doi:10.3758/BRM.40.1.61

Voss, A., Nagler, M., & Lerche, V. (2013). Diffusion models in experimental psychology: A practical introduction. *Experimental Psychology, 60,* 385–402. doi:10.1027/1618-3169/a000218

Voss, A., Rothermund, K., & Brandstädter, J. (2008). Interpreting ambiguous stimuli: Separating perceptual and judgmental biases. *Journal of Experimental Social Psychology, 44,* 1048–1056. doi:10.1016/j.jesp.2007.10.009

Voss, A., Rothermund, K., Gast, A., & Wentura, D. (2013). Cognitive processes in categorical and associative priming: A diffusion model

analysis. *Journal of Experimental Psychology: General, 142,* 536–559. doi:10.1037/a0029459

Voss, A., Rothermund, K., & Voss, J. (2004). Interpreting the parameters of the diffusion model: An empirical validation. *Memory & Cognition, 32,* 1206–1220. doi:10.3758/BF03196893

Voss, A., & Voss, J. (2007). Fast-dm: A free program for efficient diffusion model analysis. *Behavior Research Methods, 39,* 767–775. doi:10.3758/BF03192967

Voss, A., & Voss, J. (2008). A fast numerical algorithm for the estimation of diffusion model parameters. *Journal of Mathematical Psychology, 52,* 1–9. doi:10.1016/j.jmp.2007.09.005

Voss, A., Voss, J., & Klauer, K. C. (2010). Separating response-execution bias from decision bias: Arguments for an additional parameter in Ratcliff 's diffusion model. *British Journal of Mathematical and Statistical Psychology, 63,* 539–555. doi:10.1348/000711009X477581

Wagenmakers, E.-J. (2009). Methodological and empirical developments for the Ratcliff diffusion model of response times and accuracy. *European Journal of Cognitive Psychology, 21,* 641–671. doi:10.1080/09541440802205067

Wagenmakers, E.-J., van der Maas, H. L. J., Dolan, C. V., & Grasman, R. P. P. P. (2008). EZ does it! Extensions of the EZ-diffusion model. *Psychonomic Bulletin & Review, 15,* 1229–1235. doi:10.3758/PBR.15.6.1229

Wagenmakers, E.-J., van der Maas, H. L. J., & Grasman, R. P. P. P. (2007). An EZ-diffusion model for response time and accuracy. *Psychonomic Bulletin & Review, 14,* 3–22. doi:10.3758/BF03194023